DOCUMENT RESUME

ED 214 951 TM 820 093

AUTHOR Yeh, Jennie P.
TITLE Test Use in Schools. Studies in Measurement and Methodology, Work Unit 4.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Jun 78
NOTE 69p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Achievement Tests; Administrator Attitudes; Elementary Secondary Education; Interviews; National Surveys; Quantitative Tests; Questionnaires; *Teacher Attitudes; Testing Problems; *Test Use; Verbal Tests

ABSTRACT

The frequency and diversity of reading and mathematics tests, test approach (norm-referenced or criterion-referenced) and the format of measurement were studied. The contextual variables surrounding test administration and interpretation were examined. The teacher's knowledge of, and attitudes toward, testing were investigated. Results of a literature search, national survey, teacher interviews and questionnaires were illustrated. The most common form of student achievement testing was the norm-referenced standardized achievement test. Teacher developed tests were used in addition to required standardized tests. However, teachers used interactions with students and personal observations more than test results to assess students' progress. Grade level affected testing frequency. Upper grade teachers were more likely to use test results in their instructional planning. The presence of classroom aides was another variable affecting test use. Teachers with aides were less inclined to permit students to progress if they failed a test, and more inclined to provide tutoring. Teaching experience affected the use of results from norm-referenced tests. Test taking skills, test quality, and student motivation were viewed as more important factors in test scores than quality of instruction or student ability. Results indicated teachers need more information about measurement. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED2149951

DELIVERABLE: June 1978

STUDIES IN MEASUREMENT AND METHODOLOGY

WORK UNIT 4

Test Use In Schools

Jennie P. Yeh
Project Director

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

G. S. Gray

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CENTER FOR THE STUDY OF EVALUATION
Graduate School of Education
University of California - Los Angeles

June 1978

TM 820 093

"The activity which is the subject of this report was supported in part or in whole by the National Institute of Education, Department of HEW. However, the opinions expressed herein do not necessarily reflect the position or policy of the National Institute of Education, and no official endorsement by the National Institute of Education should be inferred."

TEST USE IN SCHOOLS

Introduction

Even the casual observer realizes that tests are an omnipresent part of schooling. Enormous sums of public money are expended on the assessment of children's achievement, and test publishers are advertising "new developments" at an increasing rate. What is equally clear to those who have spent time studying teaching-learning environments is that test results seem to be only infrequently used to make classroom decisions about instruction. Further, test data only rarely seem to influence policy decisions made at levels of organization above that of the classroom. Why are measurements which are taken in the name of evaluation and justified in the name of decision-making not used? Are tests employed in other, unexpected ways? What reasons can be offered to explain the fact that tests, although common elements in school operations, seem dysfunctional?

At the most general level one might suggest that school personnel do not systematically use test results because they realize what little effect this feedback has on student achievement. They shy away from the careful use of test scores for fear that these scores serve only to emphasize system failures. Their behavwor can be seen as an avoidance of punishment. Another explanation may be related to the quality of the instruments themselves. Based on NIE-supported analyses of commercially available measures, researchers at CSE have found the quality of both norm-referenced standardized achievement tests and the newer, criterion-referenced tests to be low. (See Hoepfner, 1975; Walker, 1977.)

Thus, school personnel might disregard some test results because they feel that such data cannot satisfy real decision requirements in instructional settings, especially since this information is often displayed in a form representing group effort (as in state assessment) rather than individual achievement. Since a teacher's life is obviously concerned with the successes and failures of each child in the classroom, group data would be of less interest.

Another source of inertia related to the use of test results may be teachers' lack of knowledge about the ways test design and test interpretation can influence their own decision-making. It may be a common perception that commercial test agencies "know what they're doing." The technical processes involved in test development may seem so arcane that they render teachers' personal review and judgment of tests inadequate at best. Such a view could be explained by the limited amount of information teachers receive about test development and test interpretation in their certification sequences. Their training is often limited, restricted to methods of interpreting an IQ or stanine score, with little attention given to an analysis of the advantages and limitations of constructing various types of tests. These inadequacies in teacher training are not surprising given the fact that those involved in designing and conducting that enterprise rarely have strong backgrounds in the field of measurement themselves.

Contextual variables associated with specific schools may offer one final reason for the non-use of test results. For instance, the number of instructional alternatives available in an instructional setting probably influences the amount of testing done since test results may be used as a basis for

selecting participants for enrichment or remediation programs. Another contextual variable might be the teacher's expectations of what can be learned about children from using tests. Still another contextual variable might be the presence of aides to facilitate the record keeping tasks associated with testing.

The Test Use in Schools project explored these issues of test use stimulated by the national concern with achievement of children. In order to develop descriptive bases for knowledge about the extent of testing and the uses of tests in schools, we focused attention on the following areas:

(1) the frequency and diversity of various forms of reading and mathematics measures used in schools, attending to dimensions such as the origin of the test (local, curriculum embedded or commercial); the approach used in the test (criterion-referenced, norm-referenced); the format of measurement (paper-pencil, structured performance, informal assessment).

(2) contextual variables surrounding the administration and interpretation of reading and mathematics tests (stated and understood purposes, instructional contingencies, demographic factors).

(3) teachers' knowledge or information about test construction, test interpretation, and instructional consequences of testing and their ability to apply new knowledge in hypothetical and real situations in the areas of reading and mathematics.

Multiple approaches were used in order to explore these areas, including a review of the literature, a national survey of local and state-level educational agencies, interviews with teachers and principals in California schools, and a survey of teachers' test practices and attitudes. These approaches represented phases in our investigation, with the results of our work in earlier

phases enlightening our efforts in subsequent phases. The methods and results of each approach are described in the sections that follow.

Phase 1: Literature Search

Early in the project a literature search was conducted. The following provides a brief review of the literature as it relates to the three areas of investigation mentioned above.

1. Frequency and Diversity of Various Forms of Testing. On the average, a child takes six full batteries of standardized achievement tests before he or she graduates from high school (Houts, 1975). Two hundred million achievement test forms are used each year in the United States, and this number represents only about 65% of all educational and psychological testing that is carried out. According to the Government Accounting Office (1976, SRD-76-113), 90% of the local education agencies throughout the country administer standardized, norm-referenced tests to children within their purview.

While these figures in themselves suggest the scope of the testing enterprise, one must realize that they reflect only the use of standardized norm-referenced tests. To fully comprehend the magnitude of educational testing, one must also take into account the frequency of testing with other kinds of measures--criterion referenced tests, curriculum-embedded tests, and teacher developed tests. Although there are no accurate figures reflecting the extent of use of these other tests, some studies (e.g., Baker, 1976) suggest that there is widespread use of such tests in addition to mandated use of purchased norm-referenced tests.

Reports indicate that school systems in the United States spend $24 million annually on testing elementary and secondary school students

(Mensker, 1977). One must assume that any activity so pervasive and so resource-consuming (both money and time) would have great social impact on the participants.

2. Contextual Variables Surrounding the Administration and Interpretation of Tests. The most comprehensive work specifically on teacher use of norm-referenced standardized test results was done by David A. Goslin. Two studies, -- "The Use of Standardized Ability tests in American Secondary Schools..." (1965), and "The Use of Standardized Tests in Elementary Schools..." (1965), were conducted and their results summarized in a book, Teachers and Testing (1967).

Relying on surveys in which principals and teachers were asked to state the purposes of testing at their schools, Goslin reported that in elementary schools, the ostensible purposes of tests were to diagnose individual difficulties and to disseminate information to the student. Goslin warned of the difficulty of determining how test scores were actually used in schools. He maintained that it was virtually impossible to obtain any direct information about the extent to which teachers actually used standardized test scores in the performance of their duties. According to Goslin, the differences in the amount of testing done by a school seemed to be related to the general characteristics of the school and the specific characteristics of the teachers.

The school factors that were found to effect the amount of test use were the average income level of the students' families, the personnel of the school, the characteristics of the student body, the geographical location of the school, and the per-pupil expenditure.

The amount of testing in elementary schools was positively related to the average income level of students' families, the number of full or part-time

counselors employed at the school (and the predominance of Protestants vs. Catholics in the student body). The size of the elementary school did not appear to be related to the amount of testing in the school (p. 18).

In secondary schools Goslin found that more testing was done in urban and suburban schools than in small town and rural schools. Again, school size did not appear to be related to the amount of testing. Schools with large college preparatory student bodies (and a low proportion of minority students) tended to do more testing than schools with smaller college-aspiring populations (p. 18). Although the amount of test-giving was positively related to per-pupil expenditures, the amount of test use was negatively related to per-pupil expenditures. (The fact that tests were given, then, did not necessarily mean that the results would be used.) In general, tests were used most extensively in schools having heterogeneous student populations (p. 25).

In Goslin's studies, test score reporting was also related to various school characteristics. In both elementary and secondary schools, the social class composition of the student body appeared to be related to the policies concerning the reporting of test scores (p. 27). Secondary schools with a high proportion of college bound students were more likely to provide pupils and parents with scores (p. 32). Elementary schools in upper income neighborhoods reported test scores more often to parents (p. 27). (The amount of reporting to pupils was unaffected by [redacted] level.) The amount of reporting at the secondary schools was related to [redacted] extent of testing in the school and the extent of test use (p. 32).

3. Teachers' Knowledge of Measurement and Its Implications. As might be expected, Goslin's study revealed a relationship between teacher's measurement knowledge and test use. He found that teachers who had more training in tests and measurement, more experience administering and scoring tests, and who evidenced greater familiarity with tests tended to make greater use of test scores (p. 48). However, his data suggested that elementary school teachers, a group responsible for administering most of the standardized tests given in schools, had the least formal training in measurement techniques and the least exposure to information about standardized testing (p. 36).

Based on Goslin's report on the training of teachers in tests and measurement less than 40% had more than minimal exposure, i.e., one measurement class, (both elementary and secondary) (p. 127).

As to teacher understanding of the instructional consequences of tests, a more recent study of test use is perhaps typical. The Royal Oak Study (1975), conducted by Boyd et al in a Michigan school district, using survey approaches, reported that teachers' use of test results from the district-mandated testing program was varied. There was little evidence, however, that the testing program influenced school curriculum and classroom instruction. Perhaps since there was no inservice program provided to the teachers about the meaning and interpretation of test results, many teachers did not understand the purpose of the testing program, were uncommitted to it, and were disinclined to spend time examining and making use of the results. A small number of teachers thought the test results -- (especially those from criterion-referenced tests,) were useful, to guide planning. But for the most

part, teachers felt that normed-achievement tests were selected and imposed upon teachers by administrators, and did not add new information. Test results supplied information about students' skills that was already well understood by teachers and by the students' parents. The teachers' major criticism was that the tests given were not useful for planning instruction. The author of the Royal Oak Study concluded that norm-referenced tests did not provide the kind of information the teachers desired. Another interpretation is that teachers are not equipped to determine the usefulness of this type of test results.

Goslin also looked into the possibility that teachers might be misusing tests by drilling students on actual test items (1967), but he did not find this to be a problem.

In a study involving over 400 junior high school students of less-than-average mathematics ability, Skager (1969) found that the larger the proportion of students able to answer a pretest item correctly at the beginning of the year, the more likely teachers were to rate that item as highly relevant to their planned instruction. Thus, teachers seemed inclined to teach what students already knew. Boyd's study (1975) resulted in a similar finding--that teachers did not modify their curriculum based on test results.

Goslin's study was done ten years ago. Since that time criterion-referenced measures have come into use. In our study we attempted to determine whether the shift in testing emphasis and advances in teacher training have influenced teachers' test use behavior and test reporting methods. Nothing in the literature led us to conclude that the situation had changed in the past ten years. We therefore hypothesized that a large percentage of

classroom teachers would be unsophisticated in their ability to construct tests, to select commercially prepared standardized tests, or to interpret the results of tests either for their own purposes or to convey information to students and parents.

Phase 2: National Survey

Since most state and local educational agencies respond to the demands of evaluation and accountability with programs of achievement testing, the Test Use project began by requesting information from these agencies regarding elements of their testing programs. The goal was to obtain a nation-wide picture of achievement testing.

One hundred fifteen letters requesting testing information were sent to the following people:

- Chief state school officers of the departments of education in all fifty states and the district of Columbia (51 letters)
- Directors of instructional research in the 64 nation-wide school districts with enrollments of at least 50,000 students (64 letters).

Specific information was requested in the following six categories:

1. the demographic, organizational, and resource characteristics of the schools in which tests are administered
2. the names and grade levels of the tests employed most frequently
3. the procedures by which tests are selected
4. the purposes for administering tests
5. the audiences for the test results and the methods of disseminating these results
6. the in-service or staff development provided to assist in constructing, scoring, and interpreting tests and using test results.

Replies were received from 38 chief state school officers (75%), and 35 school districts (55%). The replies ranged from personal answers to individual questions to copies of previous years' test results.

The relevant information was extracted and categorized according to the following program characteristics:

- agency requiring the testing
- names and grade levels of tests employed
- test selector
- test purposes
- test audiences
- methods of disseminating results
- in-service/staff development provided

In order to classify program characteristics answer categories were developed. For example, answers related to test purpose were placed on a continuum ranging from those having a direct, observable impact on decision-making to those having no impact on decision making. Thus an answer like "to satisfy requirements" was placed on one end of the continuum, while an answer like "to guide instructional planning" was placed on the opposite end.

In analyzing the data, state and local sources were grouped separately, and various analysis were performed (e.g., modal responses, distributions).

Care was taken to analyze the resulting information with due regard for program differences and student backgrounds. For example, the types of tests used were examined not only for their technical merit, but also for their appropriateness in terms of the age level of the target students and the purposes of the testing.

There were several problems encountered in the analysis of the data. As mentioned earlier, we contacted only large school districts, thinking that they would be most likely to have the kinds of data we were seeking. In fact, they did not always have the requested information readily available, nor were they all equally inclined to provide it. Consequently, the responses differed in their completeness and relevance.

Some questions were not answered. For example, the request for information about the characteristics of the schools in which tests were administered was ignored, possibly because of the difficulty of compiling this information.

Other questions were not interpreted uniformly by all respondents. For example, some respondents answered the question about the way test results were reported by describing the dissemination format (e.g., "by principal presentation"), while others reported the type of score dissemination (e.g., "grade-equivalent scores").

Some information was so general that it couldn't be categorized. For example, in reporting on methods of disseminating results, one response was "a summary." Another respondent stated as the purpose of their tests the identification of student strengths and weaknesses making no distinction between the local and state-mandated tests administered.

Information from publishers. Most mandated testing in schools relies upon the use of widely known published tests. We felt the publishers of these tests, since they serve so large and diverse an audience, might offer us a broader perspective on test use.

Eleven publishers whose tests are commonly used in schools were also sent letters asking for any existing information on the six topics listed above (p. 9).

Information received from test publishers added little to our data. Of the eleven publishers, 6 responded (54%). Most respondents either were not able to provide the requested information or didn't feel it was their responsibility to do so. In response to the question about the characteristics of the schools employing their tests and their purposes in using these tests, one publisher indicated that information of this type was not available. Another dismissed the questions by stating that they had no control over who used their tests or for what purposes the tests were used. A third stated that it would be a breach of their ethical responsibility to divulge such information.

Consistent with this apparently passive role, the only in-service assistance provided by most test publishers is in the form of printed material. One publisher provides a handbook about test construction. Three publishers offer printed information about test selection. Most others provide catalog information about the differences between types of tests and/or specimen sets. One publisher did have consultants available.

From this meager response, it seems that although test publishers have marketing figures which characterize their test purchasers, they do not compile data on what these purchasers do with the tests. Their responsibility seems to end with the provision of procedures for standardized test administration.

Major Findings from the National Survey

The information provided by the LEA's and SEA's can be examined in terms of testing program characteristics as well as the relationships among these characteristics. But first, some general comments can be made about the extent of achievement testing.

Nearly all respondents indicated that achievement tests were administered. The most common type of tests administered were norm-referenced tests. The most frequently used tests were the California Test of Basic Skills, the Illinois Test of Basic Skills, the Stanford Early Identification and Achievement Tests, and the Metropolitan Readiness and Achievement Tests.

In addition to these tests and the many locally developed criterion-referenced, objectives-based and objectives-referenced test, some twenty tests were mentioned by respondents. Several of this group were mentioned by just one respondent. In terms of national testing patterns there seems to be considerable latitude in test selection and seemingly little, if any, standard selection criteria.

Every grade level was tested to some degree, but the states reported relatively less testing at the lower primary levels. The local agencies, on the other hand, reported less testing at the high school level. Within these general trends, some grades were tested more than others, specifically, grades which are transition points--from lower to upper elementary, to junior high, and from junior to senior high school. The grade most often tested by the states was 8th grade, the transition point from junior to senior high (in most cases). The grade most often tested by local agencies was 6th grade, the transition point between elementary school and junior high. In general, local

agencies preferred to use the CTBS, while the state agencies preferred to use the Stanford Achievement Test (SAT).

Within the general context of norm-referenced testing at transition points, LEA's and SEA's reported differences in the characteristics of their testing programs.

Results related to the testing program characteristics

Test selector...LEA's and SEA's were asked to identify who actually selects the tests used in their testing programs since tests will invariably possess the qualities deemed important by the selectors. Both state and local agencies reported teacher input in test selection. The most common selector of both state and local tests was a committee with/without, teacher representatives. Local agencies more often reported the inclusion of teachers on selection committees than did state agencies. This use of committees can be seen as an attempt to ensure a broad representation of interests. And teacher involvement in test selection may be desirable for the following reasons:

- they are familiar with the content of instruction
- they are familiar with the student's test-taking skills
- they are the ones who generally are asked to administer the tests
- they are potentially able to make use of the results for instruction and for other purposes
- their involvement in the selection might increase their interest in the testing program and in the use of the test results.

However, it is interesting to note that even with teacher involvement on selection committees, norm-referenced tests were selected much more often than were criterion-referenced tests.

Test purpose...As mentioned earlier, respondents' answers were viewed on a continuum reflecting varying degrees of test result untilizatwon. The two purposes mentioned most frequently by state agencies were: to provide information to others (a minimal utilization response) and to evaluate instructional programs (a maximal utilization response). The two purposes mentioned least by the state agencies were maximal utilization responses normally carried out by LEA's: to plan instruction and to counsel students.

In contrast, the most common purpose mentioned by local agencies was to identify student/program strengths and weaknesses, (a relatively high utilization response); the one mentioned least often was to make student/program comparisons. As expected only local agencies mentioned the lowest utilization response, to satisfy testing requirements, since obviously the state is the initiator of these requirements.

This pattern of responses points out differences in testing purposes between state and local interests. The state performs a supervisory function by evaluating the outcomes of various educational programs and by providing this information to interested parties. It does not concern itself with student level data. These purposes are in contrast to those of local agencies which have a specific interest in obtaining student level diagnostic information for their own use.

Further insight can be gained by considering the stated purposes of both SEA's and LEA's in light of the tests they most often employ. For example,

many states mandate the Metropolitan Achievement Test for program evaluation at the 9th grade (advance level). Many LEA's use it for identifying individual student's strengths and weaknesses at the 7th and 6th grades (intermediate level). Although intermediate and advanced levels of the MAT differ in difficulty, there is considerable overlapping in the skills tested, and the method for reporting is identical. It is therefore, interesting to note that virtually the same test is used for such distinctly different purposes. Furthermore, there is some controversy as to whether a norm-referenced test such as the Metropolitan Achievement Test can perform the function of evaluating several different instructional programs within a state or assisting local sources in identifying individual student/program strengths and weaknesses. At best, the information about test purposes can indicate what these agencies want from tests rather than what they actually get.

The responses indicated a different pattern or use for criterion-referenced tests. Criterion-referenced tests were more likely to be used at several grade levels, rather than at transition points. If certain grades were singled out, testing was usually done to provide proficiency information. Testing at several grade levels is consistent with the stated purpose for using criterion-referenced tests, that is, to identify student/program strengths and weaknesses.

It is interesting to note that although there was consistency among the LEA's in terms of their stated testing purposes (to identify student/program strengths and weaknesses), there were differences in the way they sought to accomplish their purposes--some sources used criterion-referenced tests and some used norm-referenced tests.

Audiences for test results...LEA's and SEA's were asked to indicate the recipient of test results since having the results is a minimal condition

for using them. The most common audiences for state results were government representatives and school administrators. School administrators were also the ones most likely to receive local testing program results. Students were the least likely to receive either state or local test information. These findings indicate that test results do find their way back to the individual schools where they can be put to use. Students, however, are not usually considered potential users of test results in any direct way. And teachers are not the most common recipients, perhaps because of the form in which test results are reported.

The reporting of test results... If those who have the results are to make use of them, they must be reported in a comprehensible manner. Although LEA and SEA respondents did not uniformly interpret our request for information about the way results were disseminated, this much can be said: The amount and type of information provided was highly dependent upon the particular test and test publisher. Group sub-skill breakdowns were only available from certain norm-referenced tests such as the California Test of Basic Skills. Individual sub-skill breakdowns were provided for criterion-referenced tests. The Metropolitan Achievement Test was not one of the norm-referenced tests which provided group sub-skill breakdowns, and thus could not have provided state and local officials with the kinds of information needed for student/program evaluation or individual diagnosis, the purposes they stated for using the test. For the latter purpose, criterion-referenced tests are far more appropriate.

Inservice...LEA's and SEA's were asked to describe their teacher inservice training possibilities to determine how much assistance was being provided

to teachers, since presumably the more knowledge teachers have about tests, the more appropriately they will use the results of them. The responses concerning inservices were viewed on a continuum from impersonal assistance (i.e., the provision of written material) to involved assistance (i.e., workshops). Some kind of teacher inservice assistance was reported by all the respondents.

The most common types of inservice at the state level were the provision of printed material and meetings for teachers. At the local level, the most common method was the provision of printed material, although inservice training for selected staff members such as test coordinators, specialists, etc., was much more likely to be reported at the local level than at the state level.

The printed materials provided by both LEA's and SEA's were most likely to take the form of test result reports or teacher handbooks. Thus teachers had to read this information for themselves to learn about tests especially since the state seems to have left inservice training in the hands of LEA's. LEA's were more likely than were SEA's to provide test results without any interpretation to the individual schools. Perhaps this is attributable to the LEA's relatively lower level of funding for such activities. Or perhaps state agencies feel compelled to provide more complete information to individual schools because of their lack of proximity to the teachers. When schools were provided with test interpretation information, it usually consisted of an explanation of the schools' particular scores or a general explanation of the particular score reported. The least likely topics of inservice were test selection and test interpretation; topics which could provide teachers with a more general perspective on testing than specific test scores would.

Phase 3: Interviews

The information gathered from Phase 2 of the study revealed that there is great variety in the purposes of state and locally mandated tests, and that the assistance or guidance provided to teachers in terms of their interpretation and use of test scores is minimal. We therefore assumed that teachers' use of the results of such tests is, to a great extent, dependent on their individual perceptions and practices. If so, what might be the factors that determine an individual teachers' use of required test results? In addition to required tests, what techniques do they use to assess their students? Do they use tests at all? If so, with what types of tests? What factors other than state or district testing policies influence the methods a teacher chooses to assess his/her students? Do teachers from schools with more resources for student enrichment and remediation tend to test more often because they are faced with more decisions? Do teachers with large classes test more often because they have less time to spend with individual students? Do teachers use results from required tests more often if their students tend to perform well on these tests or if they regard the test as closer to the content of instruction?

In-depth teacher interviews were designed to further examine teacher test practices and their relationship to various factors in the educational process. The purposes of Phase 3 then, were:

1. to examine testing practices within as wide a range of instructional practices, school settings, and district policies as possible to identify those factors which influenced the practices.
2. to examine the ways in which test information is used by teachers in instruction.

3. to construct a model of teachers' test use behavior that would enable us to develop a questionnaire for the fourth phase of our study.

Since previous studies (Goslin, 1965, 1967) indicated that teachers testing practices varied with the location of schools and the SES of parents, we decided to select a sample of schools that covered a wide range of locations and SES groups. Five California districts--including urban, rural, and suburban schools--were selected. From these districts, schools were selected on the basis of socio-economic status and existence of outside funding (i.e., Title I, EDY, and ECE).

Within each district four elementary schools were randomly selected to create a sample of twenty schools. All 20 schools were to participate in the final phase of the study--Phase 4: Questionnaires; but only 10 of these schools were selected for the interview phase. At these 10 schools, one administrator and one teacher at each grade level were interviewed. There were 70 interviews in all.

The interviews, lasting an hour or more, were conducted by a team of 3 CSE staff members. The questions were open-ended and covered such topics as their preferred assessment methods and use of various types of test results. In addition, the teachers were asked to describe the ability level of their students, the methods and materials used for instruction and their attitude towards the tests mandated by various agencies. From the information collected during these interviews, we identified the variables that we used to develop a hypothetical model of teachers' testing practices and use of test results (see Figure 1, page 25a).

Findings from the interviews

As mentioned earlier, our assumption was that testing practices might vary because of school location (i.e., urban, rural, suburban), socio-economic status of the community (i.e., high, middle, low), and availability of instructional options (e.g., paid aides, reading specialists). Thus we selected the interview sample so as to include schools that represented the range of possibilities in the above-mentioned areas of interest.

It should also be mentioned here that is our interviews the word "tests" encompassed everything from standardized norm-referenced tests to short teacher-made quizzes. However, we did, ask our interviewees to draw a distinction between tests, work samples and teacher observation.

The information reported here does not pretend to make conclusions about differences in schools based on the above-mentioned factors, in part because of the small sample (10 schools, each representing a unique combination of the various factors) and in part because the interviews were open-ended and intended to elicit as wide a range of answers as possible.

Some impressionistic findings from these interviews will be discussed in relation to the influences on testing and instructional practices of

- parent attitudes
- student ability level
- teacher/principal attitudes and knowledge about testing.

It is interesting to note that, in general, differences in response patterns seemed to emerge most clearly between schools representing extremes of SES.

Parent attitudes toward testing

According to the teachers and principals who were interviewed, testing programs are influenced by parent attitudes toward testing. For example, staff in high SES schools reported that their parent communities were very concerned about measures of academic progress and that they also asked for comparative data, e.g., "How do our scores compare with the scores of other schools?" "How is my child doing in comparison with other 4th graders?" One administrator from an urban, high SES district, describing his community as 90% white, middle to upper class with most parents having at least a high school diploma, said that these parents wanted all the concrete, objective information they could get about student performance.

According to the staff of low SES schools, the kinds of questions asked by parents, if they asked any questions at all, tend to be much more general--e.g., "How is my child getting along?" When asked about the possibility of discontinuing mandated testing programs, teachers expressed the opinion that these parents for the most part are not even aware of the testing program and would probably not react if it were discontinued.

At the high SES schools, however, discontinuing the testing program would displease the parents, according to the staff. Even kindergarten teachers, who don't generally like to test their students, said that they would have to replace required testing with their own tests if the testing program were discontinued. At one school, the staff was attempting to re-educate the parent community who seemed anxious about standardized test scores, by down-playing the importance of such scores.

One can hypothesize that if the school staff perceive that the parent community wants test information, they will probably respond to this concern.

Student performance

Test practices and test use also seem to be influenced by students' test performance. Studies have found student test performance to be correlated with SES, the low SES schools tending to have students with low test scores on standardized tests.

And it is the staffs of these schools who were the ones indicating the greatest dissatisfaction with the appropriateness of mandated tests for their student populations and the least inclination to use the test results. Their feelings were that the students had made more progress than was indicated by the test scores. One principal from a school of this kind indicated that both staff and parents preferred the information from the school's normed criterion-referenced tests to the information obtained from the mandated testing.

One suburban school in a small district had standardized test scores lower than those of the other two schools in the district. Although the school was not a low SES school nor was it disadvantaged compared to the population of schools in California, it did have a minority of non-English speaking children whose language handicap tended to bring down the school's scores. The district had a policy which allowed parents to request the school of their choice. The staff at this particular school was concerned that parents were requesting the other schools because of their higher scores. They, (the staff) tended to devalue the testing program and reported a similar attitude on the part of their students' parents.

In contrast, at schools where students scored well, standardized test results were most often seen as having value.

Teacher/principal attitude and knowledge about testing

In general, the more knowledge a person had about testing, the more likely that person was to express the view that test results were useful. A good example of this finding was the opinion expressed by a principal from an urban, high SES district who had a Ph. D. in measurement. He felt that required testing was necessary, although he acknowledged that his teachers did not see its value in the same way as he did. He saw test scores as tools to guide professionals (not parents) and he acknowledged the need for a state evaluation program since such a program, in his view, could provide the most reliable information available. He did, however, express the opinion that when reporting results from the testing program such demographic factors as SES and transiency should be taken into consideration. According to him, a comparative score is only meaningful when the factors influencing the norm are comparable.

Another school from the same district was involved in two state funded programs which required testing. The principal at the second school felt that though some mandated testing was necessary, the amount required at his school was excessive.

Teachers from these two high SES schools expressed qualified acceptance of standardized tests. One didn't mind giving the tests if the results were reported to teachers. Another felt that a certain amount of required testing was necessary and didn't favor discontinuing it. Another didn't mind the testing, but found it something of a bother because she also tested with her own materials. In all cases, the teachers who reported using test results in planning instruction were those who had the most training and thus the best grasp of the ways in which tests results could be used.

As mentioned earlier, principals and teachers at low SES schools invariably expressed a preference for criterion-referenced rather than norm referenced tests.

Opinions expressed by the teachers from the two middle SES schools reflected a diversity in practices and attitudes toward testing. The one factor that seemed to differentiate among their testing practices was the grade level taught. Upper elementary teachers tested moreoften and placed greater reliance on test results, whereas the primary teachers seriously questioned the quality of the tests designed for their pupils. The principals of these two schools did not express strong opinions one way or another.

Phase 4: Questionnaires

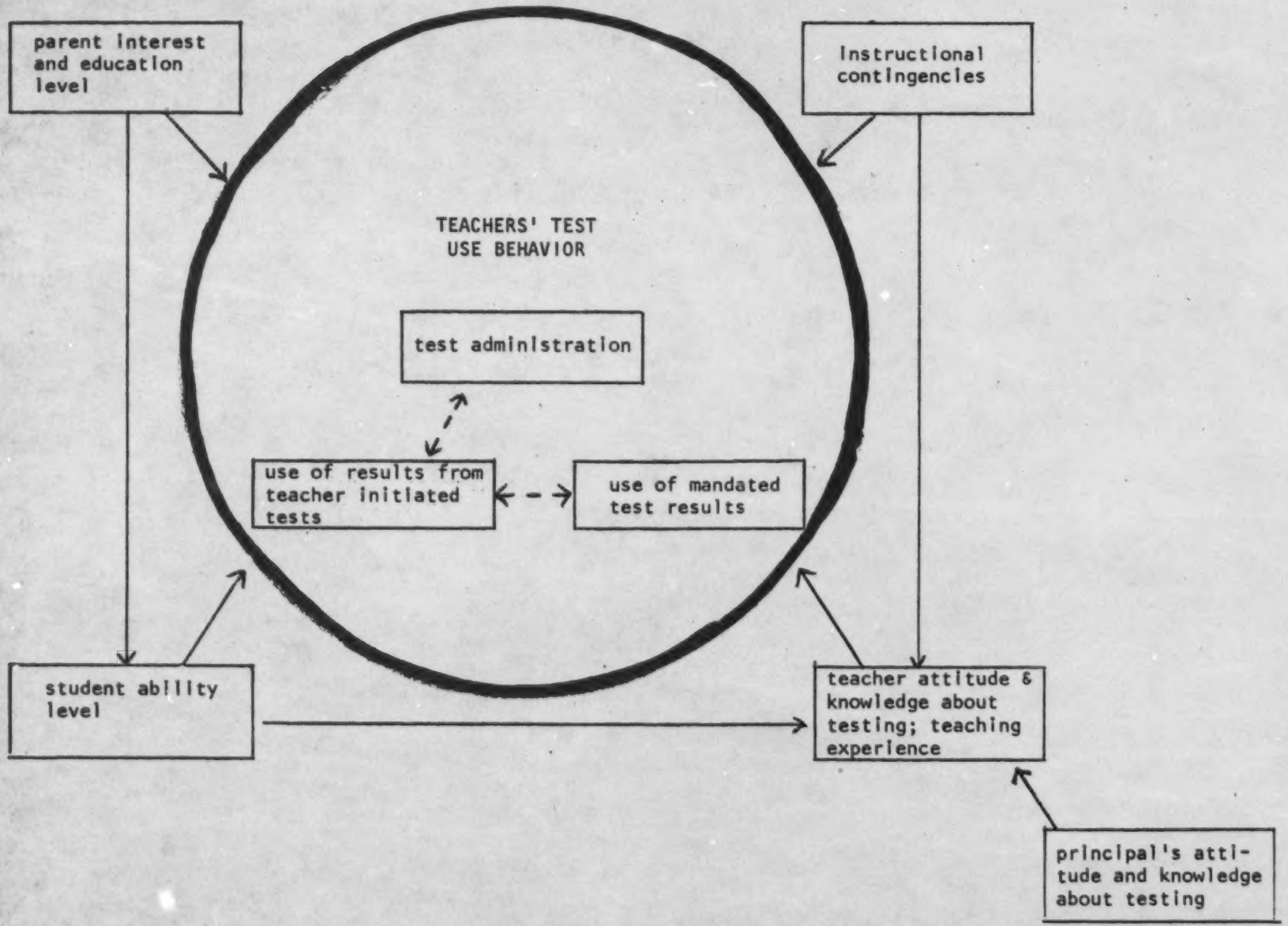
The model contained in Figure 1 was used as the basis for the development of the questionnaire. As indicated in Figure 1, teachers' test use behavior was viewed in terms of:

- test administration practices (e.g., types of tests used, frequency of testing, criteria for selecting tests)
- the use of mandated test results (e.g., how often they are used for placement of students, evaluating material and instruction, designing curriculum, communicating with parents)
- the use of results from teacher-initiated tests

Based on our review of the literature and on impressions gleaned from interviews, we hypothesized that teachers' test use behavior is influenced by the following five factors (listed in no particular order):

(1) instructional contingencies (e.g., degree of individualization availability of aides and other resources, class size, length of time a teacher spends with the same group or individual student)

FIGURE 1
Hypothytical Model of Teachers' Test Use Behavior


parent interest and education level
instructional contingencies
TEACHERS' TEST USE BEHAVIOR
test administration
use of results from teacher initiated tests
use of mandated test results
student ability level
teacher attitude & knowledge about testing; teaching experience
principal's attitude and knowledge about testing

*Arrows indicate directions of influence.

(2) students' general ability level and the dispersion of student ability within a class

(3) principal's (or administrator's) knowledge and attitude toward testing

(4) parent interest and educational level

(5) their knowledge of testing and their attitudes toward testing

The questionnaire consisted of the following sections related to the factors listed above:

Section A: classroom organization and instructional practices

B: methods teachers use to assess progress

C: use and understanding of mandated tests

D: opinions on currently used assessment methods

E: teachers' experience and training

Enough questionnaires for the estimated number of teachers at the school were mailed to principals in October 1977 and then distributed to all kindergarten through sixth grade teachers at the school. All twenty schools from the five districts received the questionnaires. However, one school was not able to participate because it was taking part in an experimental busing program that involved both students and teachers.

By the end of December 1977, 260 questionnaires were received. Even though the exact number of teachers in the 19 participating schools probably varied from the estimate obtained from the previous year's public school directory, we estimate that the return rate was approximately sixty percent.

The questionnaires were organized, coded, and then analyzed according to the hypothetical model. A componant analysis of the variance indicated that teachers within the same school tend to respond similarly to questions concerning:

•the frequency of testing and

•the types of tests used.

All other variance in testing practices is attributable to individual teachers or classrooms. Thus, analysis were performed using teacher-level data.

Major Findings from the Questionnaire Survey

The results of the questionnaire survey are discussed according to the following 15 tables which aggregate the data using various componants of the test use model:

Source of information used for initial placement (Table 1).

Teachers reported that compared to other sources of information, they most frequently used some kind of test results (from commercially published or teacher-made tests) to assess student skills in reading and math at the beginning of the school year. Fifty-eight percent reported using test results most often for initial reading placement and 65.8% reported using test results most often for initial math placement. The second most common method for student placement, informal assessment techniques (e.g., teacher observation, student work samples), was more frequently used for reading placement (24.2%) than for math placement (16.9%). The findings do not show a significant difference between the sources of information used by teachers in self-contained classrooms as compared to team teachers. Test results seem to provide the teacher with a quick and acceptable estimate of ability of those students unfamiliar to the teacher.

TABLE 1

Sources of Information Teachers Use to Place Students
In Reading and Math at the Beginning of the Year

	Reading Placement	Math Placement
A. Test results (e.g., standardized, teacher made)	58.7%(148)	65.8%(160)
B. Previous placement	15.1 (40)	16.0 (39)
C. Informal assessment (e.g., teacher observation, work sample)	24.2 (61)	16.9 (41)
D. Other	1.2 (3)	1.2 (3)

Source of information used for assessment throughout the year (Table 2).

Teachers reported that of several possible sources of information, they most frequently used information from interactions with or observations of students, informal assessment techniques (e.g., oral quizzes, reading aloud) or results from teacher-developed tests to assess their students throughout the year. The least frequently used sources of information were the results from standardized and instructional program or curriculum embedded tests, while moderate use was made of information about students' place in a book and work assignments. Apparently, once initial placements are made and teachers become more familiar with their students' ability, they are less likely to rely on standardized or instructional program tests to give them information about students' progress.

Teachers' use of these assessment methods differed with teaching experience and with the presence of an aide. Teachers with aides were more likely to use information from a variety of other sources. They were more likely to use the results from instructional program tests, students' placement in a book and the results from informal assessment techniques as well as from interaction with or observation of students. Although aides and curriculum-embedded tests may both be required by a funded program, it seems plausible that with additional help in the classroom, the teacher has more time for personal involvement with students and also more time to utilize the various sources of information available.

There were also differences in teachers' use of information sources depending on the teachers' experience level. Less experienced teachers (1-7 years of experience) were more likely than more experienced teachers (8 + years

TABLE 2

Source of Information Teachers Use To Assess Student Progress Throughout the Year

(1=never use, 5=always use)

	A standardized test results	B instructional program test results	C teacher-developed test results	D student's place in book	E work assignments	F interaction with student	G informal assessment
Any Paid Aides							
no paid aides	2.3(69)	*2.8(67)	3.9(71)	Δ3.5(72)	3.8(72)	4.2(72)	*3.8(71)
some paid aides	2.2(151)	3.2(146)	3.9(153)	3.8(150)	3.7(148)	4.3(154)	4.1(151)
Teaching Experience							
1 - 7 years	2.1(63)	2.9(60)	*4.0(64)	Δ3.9(64)	*3.9(63)	4.4(64)	Δ4.2(64)
8 - 15 years	2.4(91)	3.3(90)	3.7(92)	3.6(92)	3.7(90)	4.2(93)	4.0(91)
≥ 16 years	2.4(86)	3.1(83)	3.9(88)	3.6(86)	3.5(86)	4.2(89)	3.8(81)
Class Organization							
team-taught classroom	2.3(19)	3.3(77)	3.9(81)	3.7(79)	3.6(78)	4.2(81)	4.0(79)
self-contained classroom	2.2(134)	3.0(129)	3.8(135)	3.7(135)	3.8(134)	4.3(137)	4.0(136)
Grade Levels Taught							
K - 2nd	2.1(95)	3.2(96)	3.8(98)	3.6(95)	3.6(95)	4.2(98)	3.9(96)
3rd - 4th	2.3(48)	3.1(45)	4.0(50)	3.9(50)	3.8(49)	4.3(50)	4.0(50)
5th - 6th	2.4(58)	3.0(54)	3.9(57)	3.7(58)	3.9(57)	4.3(58)	4.1(58)
TOTAL SAMPLE	2.2(221)	3.1(215)	4.0(226)	3.7(224)	3.7(222)	4.2(227)	4.0(225)

* = one-way ANOVA significant at $\alpha \leq .05$

Δ = one-way ANOVA significant at $.05 < \alpha \leq .10$

experience) to report that they use results from their own tests, student work assignments, and informal assessment, but they showed slightly less use of results from standardized or curriculum-embedded tests. Perhaps more recent teacher training has emphasized the importance of using criterion-referenced assessment methods and pointed out the potential problems of standardized tests. But more importantly, new teachers are as likely as experienced teachers to use intuitive methods (observation of students, informal assessment) to assess student progress, but are more likely to also use concrete information (their own tests, work assignments, informal assessment).

There were no significant differences between the sources of information used by teachers teaching different grade levels. There was, however, a trend for upper-grade teachers to more frequently use standardized test results than lower grade teachers, while lower grade teachers reported more frequent use of instructional program test results.

Frequency of math and reading tests (Table 3).

In general, teachers at every grade level reported giving math tests more often than reading tests. Fifty nine percent of the teachers reported giving either monthly or weekly reading tests while 8.7% reported giving daily reading tests. This compares to 18.1% of the teachers who reported giving monthly math tests, 50% who reported giving weekly math tests, and 9.7% who reported giving daily math tests. 32.3% of the teachers gave less than 1-2 reading tests a semester, whereas 21.8% of the teachers gave less than 1-2 math tests a semester. This finding may reflect the fact that more

drill and practice is necessary in math and/or that math tests are easier to construct, score, or interpret than reading tests.

There was also a difference in teachers' testing frequency depending on the grade level taught. The results showed a trend for upper-grade teachers to give tests more often than primary grade teachers. These results seem to confirm information from the interviews that teachers gave fewer tests to younger children because they felt these students lacked the powers of concentration and skills necessary for taking tests. But in light of the previous question, when primary grade teachers did give instructional program tests, they used the results to assess student progress throughout the year. Although less experienced teachers (1-7 years) were likely to test more often than more experienced teachers, the difference was not statistically significant. Together with previous data, we get a picture of the less experienced teacher doing more assessing of students than the more experienced teacher but selectively using the results of certain types of tests. The presence of classroom aides and the type of classroom organization did not appear to effect the frequency of testing but as mentioned before, it did effect the sources of information ultimately used to assess student progress.

Types of tests administered (Table 4).

Instructional program and teacher-developed types of tests constituted the greatest proportion of all tests teachers administered to their students. This finding concerning teacher-developed tests corresponds to the previous finding that results from these tests were most frequently used to assess student progress throughout the year. Whereas a great proportion of instructional program tests were administered but least frequently used throughout

TABLE 3

How Often Teachers Give Reading and Math Tests
(0=never, 5=daily)

	A Reading Tests	B Math Tests
Any Paid Aides		
no paid aides	2.9(71)	1.2(70)
some paid aides	3.0(142)	1.3(140)
Teaching Experience		
1-7 years	3.0(64)	3.5(60)
8-15 years	3.0(84)	3.3(84)
$\geq$ 16 years	2.8(83)	3.3(83)
Class Organization		
team-taught classroom	3.0(75)	3.5(74)
self-contained classroom	3.0(133)	3.4(131)
Grade Levels Taught		
K - 2nd grade	*2.8(95)	*3.0(95)
3rd - 4th grade	3.2(48)	3.7(48)
5th - 6th grade	3.3(54)	3.9(51)
TOTAL SAMPLE	3.0(213)	3.4(212)

*one-way analysis of variance significant at $\alpha \leq .05$

TABLE 4

Teachers' Use of Different Types of Tests
(1=none, 6=all)

	A Standardized Tests	B Instructional Program Tests	C Locally Developed Tests	D Teacher-Developed Tests
Any Paid Aides				
no paid aides	2.3(63)	3.6(68)	*2.2(62)	3.5(69)
some paid aides	2.1(134)	3.6(143)	2.7(126)	3.5(134)
Teaching Experience				
1 - 7 years	2.2(59)	3.6(60)	2.6(58)	3.5(61)
8 - 15 years	2.4(78)	3.6(84)	2.6(75)	3.5(78)
$\geq$ 16 years	2.2(78)	3.5(83)	2.5(69)	3.5(80)
Classroom Organization				
team-taught classroom	2.3(71)	3.8(77)	*2.9(71)	3.6(125)
self-contained classroom	2.1(119)	3.6(127)	2.4(111)	3.6(69)
Grade Levels Taught				
K - 2nd grade	2.1(84)	*3.7(92)	2.7(82)	*3.2(82)
3rd - 4th grade	2.2(42)	4.1(47)	2.5(41)	3.7(48)
5th - 6th grade	2.2(54)	3.3(56)	2.5(49)	3.9(54)
TOTAL SAMPLE	2.2(197)	3.6(211)	2.6(188)	3.5(203)

*one-way analysis of variance significant at $\alpha \leq .05$

the year. Locally-developed tests constituted a more moderate proportion while standardized tests constituted the smallest proportion of all the types of tests teachers administered. Teachers with paid aides and teachers in team taught classrooms tended to give a larger proportion of local tests developed by someone other than themselves (e.g., district tests) than did teachers without aides and teachers in self-contained classrooms, which is only logical since these tests may have been developed by the aides or the team teachers and simply administered by the respondent teacher. Team teachers gave a greater proportion of all types of tests than did teachers in self-contained classrooms although there was no significant difference in their frequency of testing. As hypothesized earlier, the classrooms with aides may be part of funded programs which require district or school-developed tests, and this fact would account for the administration of a greater proportion of these types of tests.

Teachers of different grade levels differed in the proportion of instructional program tests they administered. Third and fourth grade teachers administered the greatest proportion of this type of test, whereas fifth and sixth grade teachers administered the lowest proportion. This finding is consistent with third and fourth grade teachers reporting of comparatively greater use of instructional program test results. Upper-grade teachers administered a greater proportion of teacher-developed tests than did primary grade teachers. Teachers at all grade levels gave an equally small proportion of standardized tests. The amount of teaching experience did not significantly affect the proportional administration of the various test types although as indicated earlier teachers with less experience more often used the results from their own tests.

Teacher development of tests (Table 5).

Fifty three percent of the teachers responded that they did develop their own tests, and 47% indicated that they did not develop their own tests. Lower grade teachers were much less likely to develop their own tests than were upper grade teachers. This finding is consistent with lower-grade teachers' general tendency to give fewer tests and their slight tendency to use instructional program tests. Teachers with more experience were slightly more likely to develop their own tests than were less experienced teachers, although this difference is not significant. It may well be that with experience, teachers gain test development skills and confidence in their own prepared material.

Reasons for developing tests (Table 6).

Teachers reported that the most important reasons for developing their own tests is that these tests more accurately assess the effects of their instruction. In other words, their own tests were seen as more content valid. Another reason considered equally important for teacher test development was that the format and wording of teacher-made tests are more suitable for students. This is especially important in the primary grades. This finding confirms the reasons teachers gave in the interviews for developing their own tests. Reasons identified as less important were lack of time to order other tests, insufficient money to buy prepared tests, and inadequate information about available tests. In other words, teachers did not develop their own tests because they felt they lacked other options. However, these findings differed by grade level taught and type of classroom. Upper grade teachers and teachers in self-contained classrooms were more likely to report that lack of time, funds, or information about prepared tests were important reasons for

TABLE 5

Do Teachers Usually Develop Their Own Reading/Math Tests

	No - do not develop tests	Yes - develops tests
Any Paid Aides		
no paid aides	40% (26)	60% (39)
some paid aides	50% (73)	50% (74)
Teaching Experience		
1 - 7 years	51% (32)	49% (31)
8 - 15 years	48% (42)	52% (45)
$\geq$ 16 years	48% (38)	52% (41)
Class Organization		
team-taught classroom	49% (38)	51% (39)
self-contained classroom	47% (61)	53% (68)
Grade Level Taught		
K - 2nd grade	63% (59)	*37% (35)
3rd - 4th grade	48% (23)	52% (25)
5th - 6th grade	30% (16)	70% (38)
TOTAL SAMPLE	47% (100)	53% (112)

*one-way analysis of variance significant at $\alpha \leq .05$

TABLE 6

Teachers' Reasons for Developing Their Own Tests

(1=very unimportant, 5=very important)

	A Insufficient funds to buy tests	B Teacher-developed tests more suitable for students	C No time to order published tests	D Information about prepared tests unavailable	E Teacher-developed tests better assess instruction
Any Paid Aides					
no paid aides	1.9(41)	4.5(42)	1.6(39)	1.9(37)	4.6(41)
some paid aides	1.7(64)	4.5(71)	1.6(65)	1.9(66)	4.6(72)
Teaching Experience					
1 - 7 years	1.6(31)	4.6(33)	1.5(31)	1.8(31)	4.7(33)
8 - 15 years	1.9(44)	4.5(45)	1.5(44)	1.9(44)	4.6(43)
$\geq$ 16 years	1.8(34)	4.5(42)	1.7(33)	1.8(33)	4.8(42)
Class Organization					
team-taught classroom	1.7(34)	4.4(33)	1.5(35)	1.7(35)	4.5(41)
self-contained classroom	1.9(68)	4.6(73)	1.7(66)	2.0(66)	4.6(70
Grade Levels Taught					
K-2nd Grade	*1.4(32)	4.4(36)	1.6(33)	*1.6(33)	4.5(37)
3rd-4th grade	2.1(22)	4.7(24)	1.7(21)	2.5(21)	4.7(23)
5th-6th grade	2.1(37)	4.5(39)	1.8(36)	2.1(36)	4.5(39)
TOTAL SAMPLE	1.8(109)	4.5(113)	1.6(104)	1.9(103)	4.6(113)

*one-way analysis of variance significant at $\alpha \leq .05$

developing their own tests. Perhaps since they were more likely to feel that prepared tests were suitable for their students, they were more aware of the constraints inhibiting them. There was no difference in responses reported by teachers with and without aides, and teachers with little or lengthy experience. Its seems that these groups of teachers were equally concerned with providing appropriate tests for their students.

Reasons for not developing tests (Table 7).

Teachers reported that the most important reason for not developing their own tests was that suitable published tests were available. Teachers reported that the reason least important for not developing their own tests was that they do not test their students unless it is required by the school or other outside agencies.

Teachers with aides were significantly more likely to report that suitable published tests were available, again reflecting the aide-funded program/testing combination. Less experienced teachers were significantly more likely to report that lack of time was an important reason for not developing their own tests. It is common sense that people with less time on the job are less efficient at managing their time. There were no significant differences between teachers in self-contained vs. team-taught classroomw, although team teachers considered time, training in test construction, and importance of student norms as less of a deterrent to self-test development than other teachers. Perhaps time is less of deterrent because teaching duties are shared. There was a trend from lower to upper-grade teachers in their desire for normative data, with upper-grade teachers more likely to indicate the need for national or local comparison information as an important reason for not developing their own tests.

TABLE : 7

Teachers' Reasons for Not Developing Their Own Tests

(1=very unimportant, 5=very important)

	A Comparison of students to national/local norms important	B Suitable published tests un-available	C No time to for teacher-developed tests	D Testing only done when required	E More teacher training in test construction necessary
Any Paid Aides					
no paid aides	2.8(44)	$^{\Delta}$3.4(45)	2.4(41)	2.5(42)	2.5(41)
some paid aides	2.7(86)	3.8(89)	2.4(83)	2.1(79)	2.4(80)
Teaching Experience					
1 - 7 years	2.8(40)	3.5(40)	*2.8(41)	2.2(40)	2.2(41)
8 - 15 years	2.8(54)	3.8(56)	2.3(53)	2.1(49)	2.8(51)
$\geq$ - 16 years	2.6(49)	3.8(51)	2.1(43)	2.6(45)	2.3(43)
Class Organization					
team-taught classroom	2.7(47)	3.7(49)	2.2(45)	2.3(44)	2.3(43)
self-contained classroom	2.9(83)	3.7(86)	2.6(80)	2.3(77)	2.6(78)
Grade Levels Taught					
K - 2nd grade	2.6(64)	3.8(66)	2.3(62)	2.4(57)	2.5(60)
3rd - 4th grade	2.8(28)	3.7(28)	2.8(26)	2.3(27)	2.6(25)
5th - 6th grade	3.1(31)	3.5(32)	2.2(29)	1.9(29)	2.4(29)
TOTAL SAMPLE	2.7(130)	3.7(134)	2.4(124)	2.3(121)	2.5(121)

*one-way analysis of variance significant at $\alpha \leq .05$

$^{\Delta}$one-way analysis of variance significant at $.05 < \alpha \leq .10$

This finding confirmed the survey data from Phase 2 which showed a trend for norm-referenced testing at transition points (grades 5-6). There was also a tendency for the lower-grade teachers to cite testing only when required as a deterrent to developing their own tests more often the upper grade teachers. This would indicate that lower-grade teachers test less often than upper-grade teachers and tend to use already prepared tests. This finding was substantiated in earlier sections.

Qualities of prepared tests (Table 8).

The qualities teachers considered most important when choosing prepared tests were clear format, similarity to class material, and accurate prediction of achievement. These qualities were the same as the reasons teachers gave for developing their own tests. The quality teachers considered least important was the test publisher's good reputation. There was a significant difference in responses towards test format, with kindergarten-2nd grade teachers more likely to rate format as a very important quality than upper grade teachers.

There was also a trend for lower grade teachers to consider the ease of test administration and scoring as a very important quality. This finding confirms primary grade teachers' concern that younger students' lack of test taking skills may affect the test results, unless the test has a clear format. The presence of aides, classroom from organization, and teaching experience showed no correlation with teachers' choices of important test attributes. Again suggesting that these groups share a common desire for valid and reliable tests.

TABLE 8

Qualities That Teachers Consider When Selecting Tests

(1=very unimportant, 5=very important)

	A Test material similar to that presented in class	B Test simple to administer and/or score	C Test has clear format, directions	D Test publisher has a good reputation	E Test accurately predicts student achievement
Any Paid Aides					
no paid aides	4.4(70)	4.2(69)	4.5(70)	2.6(68)	4.4(70)
some paid aides	4.6(149)	4.2(150)	4.7(150)	2.9(146)	4.4(149)
Teaching Experience					
1 - 7 years	4.4(65)	4.1(65)	4.5(65)	2.7(64)	4.4(65)
8 - 15 years	4.6(89)	4.2(90)	4.7(90)	2.7(88)	4.4(91)
$\geq$ - 16 years	4.5(84)	4.2(83)	4.6(84)	2.9(79)	4.5(82)
Class Organization					
team-taught classroom	4.6(80)	4.2(79)	4.6(79)	2.7(76)	4.3(79)
self-contained classroom	4.5(140)	4.3(141)	4.6(142)	2.8(138)	4.4(141)
Grade Levels Taught					
K - 2nd grade	4.6(96)	4.3(96)	*4.7(95)	2.8(91)	4.5(94)
3rd - 4th grade	4.7(50)	4.2(50)	4.6(50)	2.9(48)	4.5(50)
5th - 6th grade	4.3(56)	4.0(56)	4.4(57)	2.9(56)	4.3(56)
TOTAL SAMPLE	4.5(219)	4.2(219)	4.6(220)	2.8(214)	4.4(219)

*one-way analysis of variance significant at $\alpha \leq .05$

Use of teacher developed test results (Table 9).

Not surprisingly, teachers indicated that they used results from their own tests frequently or always for all of the listed purposes (instructional decisions, evaluation of methods, grade assignment, information to others). Their utility for such purposes may explain why they were one of the most frequently used assessment methods. Teachers most frequently used information from their own tests for making instructional decisions about students, for evaluating the effectiveness of teaching methods and materials, and for providing information to parents and/or staff. Their own test information was used least often for assigning grades, although it was still used between sometimes and frequently. (It may be that teachers base their grades on information other than test results.)

Responses differed depending on the presence or absence of aides, grade level taught, and teachers' level of experience. Teachers without aides tended to report using test results for grading more often than did teachers with aides. This may be explained by the fact that teachers without aides have less time and thus less opportunity to obtain information other than from test results for grading purposes. They may perceive group testing as the quickest and most efficient way of obtaining information for grading purposes. There was also a significant difference between upper and lower grade teachers' responses, in that lower grade teachers were less likely to use test results to assign grades than were upper grade teachers. This finding is consistent with lower grade teachers' lack of emphasis on testing and probably lesser emphasis on grades. Less experienced teachers, teachers with no aides, and team tea hers were more likely to use results from their own tests to provide information to parents and staff than teachers in their respective counterpart groups.

TABLE 9

How Teachers Use Information From Their Own Tests
(1=never, 4=always)

	A To make instructional decisions	B To evaluate effectiveness of methods/ materials	C To assign grades	D To provide information to parents/ staff
Any Paid Aides				
no paid aides	3.3(68)	3.2(68)	$^{\Delta}$2.8(67)	3.1(67)
some paid aides	3.2(147)	3.1(147)	2.5(145)	2.9(146)
Teaching Experience				
1 - 7 years	3.3(64)	3.2(64)	2.5(63)	3.1(63)
8 - 15 years	3.1(89)	2.9(89)	2.5(87)	2.9(89)
$\geq$ 16 years	3.2(82)	3.1(83)	2.6(80)	2.9(82)
Class Organization				
team-taught classroom	3.3(78)	3.1(79)	2.7(77)	*3.1(75)
self-contained classroom	3.2(138)	3.1(137)	2.5(136)	2.9(138)
Grade Levels Taught				
K - 2nd grade	3.2(90)	3.0(91)	*2.4(88)	2.9(89)
3rd - 4th grade	3.2(50)	3.2(50)	3.0(49)	3.0(50)
5th - 6th grade	3.2(55)	3.2(55)	2.8(56)	3.1(55)
TOTAL SAMPLE	3.2(215)	3.1(215)	2.6(212)	3.0(213)

*one-way analysis of variance significant at $\alpha < .05$

$^{\Delta}$one-way analysis of variance significant at $.05 < \alpha \leq .10$

Again, it may be that teachers with less opportunity for student teacher interaction and teachers who are less confident in their own judgments may be those likely to depend on test results to communicate with others. Although team teachers were more likely than teachers in self-contained classes to use their own test results for making instructional decisions and assigning grades, the difference was not significant.

Teachers' responses to students' failing a test (Table 10).

The most frequent action that teachers reported taking when students failed a test was to reteach the same concept or skill. The least frequent teacher action was to allow the student to progress. Responses differed significantly between teachers who had aides and those who did not. If a teacher did not have an aide, s/he was less likely to reteach, provide in-class tutoring, or assign additional practice with different materials, but was more likely to let a student progress to the next unit. The presence of an aide appears to free the teacher for these remedial types of activities and/or provides another adult to work with students who need extra help. Compared to teachers with more than 16 years of teaching experience, teachers with 1-7 years teaching experience were significantly ($\alpha \leq .10$) more likely to assign additional practice with different materials. They were also more likely to provide in-class tutoring and assign additional practice with the same materials (although the difference was not statistically significant). Perhaps newer teachers try many methods until they find one that succeeds. Lower grade teachers were significantly more likely than upper grade teachers to reteach the same concept or skill, but this result may be due to the higher

TABLE 10

What Teachers Do When Students Fail a Test

(1=most frequent, 6=least frequent)

	A Reteach concept/ skill	B Let student progress	C Assign additional practice with same materials	D Provide in-class tutoring	E Assign additional practice with different materials
Any Paid Aides					
no paid aides	*1.7(72)	$^{\Delta}$4.8(67)	2.6(69)	$^{\Delta}$2.5(69)	*2.5(69)
some paid aides	1.3(153)	5.1(148)	2.6(149)	2.1(153)	2.0(152)
Teaching Experience					
1 - 7 years	1.6(65)	5.0(65)	2.5(65)	2.3(65)	$^{\Delta}$1.9(65)
8 - 15 years	1.3(92)	5.2(89)	2.7(89)	2.1(91)	2.1(92)
$\geq$ - 16 years	1.4(87)	5.0(80)	2.7(83)	2.5(85)	2.4(84)
Class Organization					
team-taught classroom	1.3(81)	5.1(78)	2.8(78)	2.2(79)	1.9(69)
self-contained classroom	1.5(143)	5.0(138)	2.5(141)	2.3(143)	2.1(142)
Grade Levels Taught					
K - 2nd grade	*1.2(97)	5.1(92)	2.5(95)	2.2(97)	2.1(96)
3rd - 4th grade	1.5(50)	5.2(50)	2.7(50)	2.4(50)	2.1(49)
5th - 6th grade	1.5(57)	4.9(54)	2.7(55)	2.3(56)	2.3(56)
TOTAL SAMPLE	1.4(225)	5.1(215)	2.6(218)	2.3(222)	2.2(221)

*one-way analysis of variance significant at $\alpha \leq .05$

$^{\Delta}$one-way analysis of variance significant at $.05 < \alpha \leq .10$

incidence of aides in lower grade classrooms. Another possibility is that the material covered in these grades must be mastered to move on to more difficult concepts. Although the response differences were not significant, team-teachers (like teachers with aides) were more likely to provide some type of remedial activities than teachers in self-contained classrooms, again because there is more adult help for slower students.

Use of required test information (Table 11).

Teachers most often used required test information for reporting to parents and/or staff and for evaluating the effectiveness of teaching methods and materials. Required test results seemed to function as a standard of comparison. This result can be compared to use teachers make of their own test results which are more for instructional decisions. Teachers without aides were more likely than teachers with aides to use required test results to provide information to parents, again because lesser individual student contact may influence reliance on more objective types of information.

More experienced teachers were more likely than less experienced teachers to use results from required tests for evaluating the effectiveness of teaching methods and materials, assigning grades, and providing information to parents and staff. Perhaps newly trained teachers are more aware of the problems involved in using norm-referenced tests for a variety of purposes or more experienced teachers feel more familiar with using this kind of test information. There was also a significant response difference between self-contained and team teachers, team teachers being more likely to use required test results for evaluating the effectiveness of teaching methods and providing information to parents, again because the increased number of students reduces the

TABLE 11

How Teachers Use Information From Required Tests

(1=never, 4=always)

	A To make instructional decisions	B To evaluate effectiveness of methods/ materials	C To assign grades	D To provide information to parents/ staff
Any Paid Aides				
no paid aides	1.9(67)	2.0(67)	1.3(67)	*2.3(68)
some paid aides	1.8(143)	1.9(143)	1.4(140)	2.0(139)
Teaching Experience				
1 - 7 years	1.7(60)	*1.8(60)	*1.2(60)	$^{\Delta}$1.9(60)
8 - 15 years	1.9(88)	1.9(88)	1.4(87)	2.1(87)
$\geq$ 16 years	1.9(81)	2.1(81)	1.4(79)	2.2(78)
Class Organization				
team-taught classroom	1.9(69)	*2.2(69)	1.4(67)	*2.3(68)
self-contained classroom	1.8(139)	1.9(139)	1.4(139)	2.0(138)
Grade Levels Taught				
K - 2nd grade	1.8(86)	1.9(86)	1.4(86)	2.1(86)
3rd - 4th grade	1.9(48)	1.9(48)	1.4(47)	2.0(46)
5th - 6th grade	1.8(57)	2.1(57)	1.3(55)	2.2(56)
TOTAL SAMPLE	1.8(210)	2.0(210)	1.4(207)	2.1(207)

*one-way analysis of variance significant at $\alpha < .05$

$^{\Delta}$one-way analysis of variance significant at $.05 < \alpha \leq .10$

opportunity for obtaining information through interactions or observation. However, there was no significant difference between these two groups in their use of required test results for assigning grades reflecting their consensus that this information was inappropriate for that particular purpose. The grade level taught did not appear to have an influence on teachers' responses in this category.

Interpreting test results

Only 50% of the teachers were able to correctly interpret percentiles and grade equivalent scores (the two most common methods used for reporting test scores) and describe the factors that influence test performance. Thus 50% did not understand common methods used in reporting test results or factors that must be kept in mind when interpreting these results.

Factors influencing required test results (Table 12).

Teachers felt that the factors having the greatest influence on student scores on required tests were students' test taking skills, student motivation, the quality of the test, and unusual circumstances (e.g., distractions, special activities scheduled). Parental influence was reported as having some influence on re-uired test scores as did students' true ability, IQ, and quality of instruction. Responses differed according to teacher experience. Less experienced teachers perceived quality of instruction and IQ as having less influence on test scores than did more experienced teachers. Perhaps new teachers are not as sure of their influence, or perhaps these teachers may be more skeptical about the tests' sensitivity to their instruction. Although not significant, there was a trend for lower grade teachers

TABLE 12

Teachers' Opinions of Factors Which May Influence Student Test Scores on Required Tests
(1=little influence, 5=great influence)

	A Students' test-taking skills	B Quality of classroom instruction	C Student IQ	D Student motivation	E Parent interest	F Quality of test	G Unusual circum-stances	H Students' true ability
Any Paid Aides								
no paid aides	4.3(73)	3.7(72)	3.6(73)	4.3(72)	3.3(70)	4.4(72)	4.2(73)	3.8(73)
some paid aides	4.5(151)	3.7(148)	3.7(152)	4.2(152)	3.3(149)	4.4(149)	4.2(149)	4.0(151)
Teaching Experience								
1 - 7 years	4.4(65)	*3.5(64)	*3.5(65)	4.2(65)	3.3(65)	4.4(64)	4.2(65)	3.8(65)
8 - 15 years	4.5(92)	3.6(90)	3.8(92)	4.2(92)	3.1(88)	4.4(91)	4.1(91)	3.9(92)
$\geq$ - 16 years	4.4(88)	3.9(87)	3.9(89)	4.4(88)	3.4(87)	4.3(86)	4.2(87)	4.1(87)
Class Organization								
team-taught classroom	4.4(81)	3.8(78)	3.8(81)	4.3(80)	3.4(78)	4.4(79)	4.2(80)	*4.2(80)
self-contained classroom	4.5(142)	3.7(142)	3.7(143)	4.3(143)	3.2(140)	4.4(141)	4.2(142)	3.9(143)
Grade Levels Taught								
K - 2nd grade	4.4(97)	3.7(93)	$^{\Delta}$3.8(97)	4.2(96)	3.2(95)	4.4(95)	4.1(96)	4.0(97)
3rd - 4th grade	4.3(50)	3.7(49)	4.0(50)	4.2(50)	3.2(49)	4.3(49)	4.2(49)	3.9(50)
5th - 6th grade	4.5(57)	3.7(58)	3.5(58)	4.4(58)	3.4(55)	4.4(56)	4.3(57)	3.8(57)
TOTAL SAMPLE	4.4(224)	3.7(220)	3.7(225)	4.3(224)	3.3(219)	4.4(221)	4.2(222)	3.9(224)

*one-way analysis of variance significant at $\alpha \leq .05$

$^{\Delta}$one-way analysis of variance significant at $.05 < \alpha \leq .10$

to more highly rate students' IQ and true ability as influencing test scores, while upper-grade teachers considered students' test taking skills and unusual circumstances as more influential. Perhaps this can be traced to a difference in test content; with tests for lower grades heavily concerned with basic skills while upper grade tests are more influenced by individual school curricula. There was no difference in responses between teachers with or without paid aides. Team teachers felt that scores were more influenced by students' true ability than teachers in self-contained classrooms.

Influence of classroom instruction on required test scores (Table 13).

This question was intended to corroborate responses to the previous question. The results confirmed the responses of the less experienced teachers. Teachers with more than 16 years experience perceived classroom instruction as having a greater influence on scores than did teachers with less experience. Perhaps teachers become confident of their influence over the years, or just gain more confidence in required testing.

Cost benefit of required testing (Table 14).

When considering the time spent in required testing vs the benefits that accrued, teachers felt that the amount was between "about right" and "slightly too much." Responses differed significantly depending on teachers' experience level. Teachers with less than 16 years experience were more likely to respond that too much time was spent in testing in relation to the benefits received than were teachers with at least 16 years of experience. This finding may reflect less experienced teachers' dissatisfaction with the actual tests or more experienced teachers' acceptance of that testing program. Teachers with 16 years or more experience indicated that the cost-benefit ratio of required testing was appropriate.

TABLE 13

Teachers' Opinion of the Influence of Classroom Instruction on Students' Scores on Required Tests

(1=little effect, 5=great effect)

Any Paid Aides	
no paid aides	3.9 (73)
some paid aides	3.8 (153)
Teaching Experience	
1 - 7 years	*3.7 (65)
8 - 15 years	3.7 (92)
$\geq$ - 16 years	4.1 (90)
Class Organization	
team-taught classroom	3.9 (81)
self-contained classroom	3.9 (144)
Grade Levels Taught	
K - 2nd grade	3.9 (98)
3rd - 4th grade	3.7 (50)
5th - 6th grade	3.9 (58)
TOTAL SAMPLE	3.9 (226)

*one-way analysis of variance significant at $\alpha \leq .05$

TABLE 14

Teachers' Opinion About The Amount of Classroom Time Spent in Required Test Taking in Relation to Teacher/Student Benefirs Received (1=too much, 5=not enough)

Any Paid Aides	
no paid aides	2.3 (71)
some paid aides	2.3(152)
Teaching Experience	
1 - 7 years	*2.2 (64)
8 - 15 years	2.2 (93)
≥ 16 years	2.6 (86)
Class Organization	
team taught classroom	2.4 (80)
self-contained classroom	2.3(141)
Grade Levels Taught	
K - 2nd	2.3 (96)
3rd - 4th	2.3 (50)
5th - 6th	2.5 (57)
Total Sample	2.3(223)

* = one-way ANOVA significant at x<=.05

Reactions to discontinuation of required testing (Table 15).

Teachers generally perceived that other teachers would react favorably to the discontinuation of testing, that administrators would have neutral feelings, and that parents' reactions would range from neutral to negative. Teachers' responses differed depending on the presence or absence of an aide, and teacher experience level. Teachers with aides reacted more positively to discontinuing testing than teachers without aides. Again, teachers with aides may have required program tests which represent additional time in testing. Teachers with 8-15 years experience and teachers in self-contained classrooms reacted more positively to discontinuing testing than did their respective counterparts. Teachers with that level of experience also felt that the testing took too much time. Required testing may represent an additional burden to the teacher without other adult help. Less experienced teachers were more likely to perceive principals as having a negative reaction towards discontinuation of testing than were more experienced teachers. Upper grade teachers were more likely to rate parents' and administrators' reactions as negative than were lower grade teachers, perhaps because they do more required testing at those grade levels.

TABLE 15

Teacher Opinion on Reactions Should Required Testing Be Discontinued
(1-negatively, 5=positively)

	Parents' reaction	Teachers' reaction	Administrators' reaction
Any Paid Aides			
no paid aides	2.4(72)	*3.5(72)	2.7(70)
some paid aides	2.5(153)	4.0(151)	3.0(145)
Teaching Experience			
1 - 7 years	2.5(64)	3.7(64)	$^{\Delta}$2.5(62)
8 - 15 years	2.5(92)	4.0(91)	3.0(88)
$\geq$ 16 years	2.5(89)	3.7(88)	3.0(85)
Classroom Experience			
team-taught classroom	2.5(80)	3.6(80)	2.8(77)
self-contained classroom	2.6(144)	3.9(143)	2.9(137)
Grade Levels Taught			
K - 2nd grade	2.6(98)	*4.0(98)	3.0(94)
3rd - 4th grade	2.5(50)	3.8(49)	3.1(48)
5th - 6th grade	2.3(58)	3.5(58)	2.8(57)
TOTAL SAMPLE	2.5(225)	3.8(223)	2.9(215)

*one-way analysis of variance significant at $\alpha \leq .05$

$^{\Delta}$one-way analysis of variance significant at $.05 < \alpha \leq .10$

Conclusion

Our study sought to investigate three major areas of test use in schools. (1) the frequency and types of tests administered in schools; (2) the major contextual variables affecting test use; and (3) teachers' knowledge of and attitudes toward testing. Our major findings with regard to each of these areas are summarized below.

With regard to frequency of testing, the results of both our national surveys of LEA's and SEA's, as well as those from our teacher interviews, validated that testing is a fact of life in schools. Almost all the SEA's and LEA's surveyed indicated that they required some form of student achievement testing, the most common of which was norm-referenced standardized achievement tests. Teachers, too, reported using many tests, both required and non-required. In contrast to the LEA's and SEA's, teachers reported frequent use of program embedded tests and teacher developed tests. However, despite the frequency of testing, they reported that they were more likely to use interactions with students and personal observations than test results to assess students' progress.

With regard to contextual variables, grade level was one factor that affected testing frequency. Our investigation indicated that although all grade levels were tested to some degree, there is less testing at the primary level. These results were consistent for our LEA survey as well as our school studies. Supporting this finding, data from our teacher survey indicated that upper grade teachers were also more likely to use test results in their instructional planning.

The presence of classroom aides was another variable that appeared to affect test use. Teachers who had aides were more likely to use test results. Perhaps this result is a function of the instructional alternatives available to teachers with this kind of assistance. Our results indicated that teachers with aides were less likely to allow students to progress if they failed a test, and more likely to provide students with tutoring or additional practice than teachers who did not have aides.

Years of experience in teaching was still another variable that was found to affect test use. Teachers with more teaching experience were more likely to use resultsfrom norm-referenced tests than less experienced ones.

With regard to teachers' attitudes and knowledge about testing, we found that teachers do not perceive that they have much affect on their students' performance on required tests. They indicated that test taking skills, test quality, and student motivation were more important factors in students' test scores than were quality of instruction or student ability.

Our results also indicated that teachers need more knowledge about measurement. Although a majority of teachers reported having some kind of training in testing, only about 50% could accurately interpret percentile scores or grade equivalent scores--two rudimentary testing concepts.

Given these findings about teachers' knowledge and the fact that teachers indicated they wanted more training on how to use and construct criterion-referenced tests, it may be that teachers need more training before any potential value of test use is realized.

REFERENCES

Boyd, J., Jacobsen, K., McKenna, B. H., Stake, R. E., and Yashinsky, J. A Study of Testing Practices in the Royal Oak (Michigan) Public Schools. Michigan: Royal Oak Michigan School District, 1975.

Goslin, D. A. Teachers and Testing. New York: Russell Sage Foundation, 1967.

Goslin, D. A., Epstein, R., and Hallock, B. A. The Use of Standardized Tests in Elementary Schools. Second Technical Report. New York: Russell Sage Foundation, 1965.

Goslin, D. A. The Use of Standardized Ability Tests in American Secondary Schools and their Impact on Students, Teachers, and Administrators. New York: Russell Sage Foundation, 1965.

Hoepfner, R., Bradley, P. A., and Doherty, W. J. "National Priorities for Elementary Education." CSE Monograph Series in Evaluation, No. 2, 1973.

Houts, P. L. "Standardized Testing in America, II." The National Elementary Principal, 54, 6, p. 2-3.

Lortie, D. C. School-Teacher - A Sociological Study. Chicago: The University of Chicago Press.

Skager, R. W. "Student Entry Skills and Evaluation of Instruction Programs: A Case Study." A paper presented at the Annual Meeting of the American Educational Research Association, 1969.

Walker, C. B. Criterion-Referenced Test Evaluation Book, 1977, in press.